Title

Analysis of threats to research validity introduced by audio recording clinic visits: Selection bias, Hawthorne effect, both, or neither?

Authors

Henry, Stephen G

Jerant, Anthony

Iosif, Ana-Maria

et al.

Peer reviewed

Analysis of threats to research validity introduced by audio recording clinic visits: Selection bias, Hawthorne effect, both, or neither?

Stephen G. Henry[a,b], **Anthony Jerant**[b,c], **Ana-Maria Iosif**[d], **Mitchell D. Feldman**[e], **Camille Cipri**[b], and **Richard L. Kravitz**[a,b]

[a]Department of Internal Medicine, University of California Davis, Sacramento, California, USA

[b]Center for Healthcare Policy and Research, University of California Davis, Sacramento, California, USA

[c]Department of Family and Community Medicine, University of California Davis, Sacramento, California, USA

[d]Department of Public Health Sciences, University of California Davis, Davis, California, USA

[e]Department of Medicine, University of California San Francisco, San Francisco, California, USA

Abstract

Objective—To identify factors associated with participant consent to record visits; to estimate effects of recording on patient-clinician interactions

Methods—Secondary analysis of data from a randomized trial studying communication about depression; participants were asked for optional consent to audio record study visits. Multiple logistic regression was used to model likelihood of patient and clinician consent. Multivariable regression and propensity score analyses were used to estimate effects of audio recording on 6 dependent variables: discussion of depressive symptoms, preventive health, and depression diagnosis; depression treatment recommendations; visit length; visit difficulty.

Results—Of 867 visits involving 135 primary care clinicians, 39% were recorded. For clinicians, only working in academic settings (*P*=0.003) and having worked longer at their current practice (*P*=0.02) were associated with increased likelihood of consent. For patients, white race (*P*=0.002) and diabetes (*P*=0.03) were associated with increased likelihood of consent. Neither multivariable

DECLARATION:
I confirm that all personal identifiers have been removed or disguised so the persons described are not identifiable and cannot be identified through the details of the story.

CONFLICTS OF INTEREST:
None to disclose.

regression nor propensity score analyses revealed any significant effects of recording on the variables examined.

Conclusion—Few clinician or patient characteristics were significantly associated with consent. Audio recording had no significant effect on any dependent variables.

Practice Implications—Benefits of recording clinic visits likely outweigh the risks of bias in this setting.

Keywords

audio recording; Hawthorne effect; selection bias; consent; primary care; depression

1. Introduction

Patient-clinician communication is central to accurate diagnosis and effective management of common primary care problems, including physical and mental health conditions and issues such as treatment adherence.[1–3] Studying communication during primary care visits is thus important for evaluating and improving the quality and appropriateness of care. Direct observation of patient-clinician interactions via audio or video recording is one useful method of assessment,[4, 5] but we know little about the extent to which audio recording may affect the internal or external validity of research or quality improvement studies. **Studies involving recording have validity (i.e., are unbiased) if two otherwise identical clinic visits, differing only in whether the participants consented to and/or underwent recording, produce similar results**.

A minority of clinicians and patients are uncomfortable with the prospect of being recorded, [6, 7] prompting researchers to worry that recording may introduce selection bias[8] that reduces studies' validity. Published consent rates in studies that involve recorded primary care visits range from 7 to 100% for clinicians and from 17 to 100% for patients.[9] Only one prior study examined clinician consent to recording; clinicians who disapproved of video recording or felt pressed for time were less likely to consent.[10] Six prior studies (only one of which was published in the last 10 years) have examined patient factors associated with consent.[6, 11–15] Concerns about confidentiality[6] and sensitive topics (e.g., mental[11, 12, 15] or sexual health[6, 12]) were the only consistent predictors of patient consent.

In addition to introducing selection bias, recording might influence patient-clinician interactions in ways that affect the processes or outcomes under study (i.e., a Hawthorne effect).[16, 17] In one prior study, being recorded was not associated with patient satisfaction.[18] The persistent suspicion that audio- or video-recording alters clinician and patient behavior[12, 19–21] and the paucity of empirical studies addressing this question underscore the need for further research.

To address these issues, **we performed secondary analysis of data from a previously published clinical trial** designed to engage primary care patients in communication about depression.[22] Participating patients and clinicians were asked to provide optional consent to have their visits audio recorded. We identified clinician and patient characteristics

associated with consent to be recorded and then analyzed differences in communication about depression, communication about preventive health, treatment recommendations, and clinician-reported visit burden (operationalized as clinician-reported visit difficulty and visit length) between recorded and non-recorded visits. We focused on communication about depression because prior work indicates that effects of recording, if they exist, are likely to be more pronounced during visits involving discussion of mental health problems.[11, 12, 15]

2. Methods

2.1 Recruitment and randomization

Data were **from** a clinical trial comparing two patient interventions to increase discussion of depressive symptoms (an interactive computer program and a depression engagement video) and a control (a sleep hygiene video). Primary care clinicians and adult patients were recruited from several sites in either Sacramento or San Francisco, California (two Veterans Affairs (VA) primary care clinics, two academic primary care practices, two multispecialty group practices, one health maintenance organization, and one urgent care clinic). Patient inclusion criteria included the ability to speak English, the ability to use a computer, and having an appointment with a participating clinician. Patients taking medications for depression were excluded.

Patients were told that the study purpose was to improve communication about common health problems and were screened for eligibility and depressive symptoms with the Patient Health Questionnaire (PHQ)-8[23] 1–2 weeks before their visit. Patients in the urgent care clinic were screened immediately before their visit. Patients with depressive symptoms were oversampled. Non-depressed patients were included to evaluate the possibility of intervention-induced overtreatment. Full details of the parent study, which was approved by the institutional review boards of all participating institutions, have been published.[22, 24]

Immediately before their appointment, patients gave written informed consent and completed a computer-based questionnaire. Patients were then randomized to receive one of the two interventions or the control prior to their appointment. Clinicians were told that the study goal was to improve communication about overlapping mental and physical health problems and were blind to patient randomization assignment.

During the informed consent process, both clinicians and patients were asked for permission to audio record their study visits. This additional consent was optional; **recording was not part of randomization and** did not affect other aspects of study participation **in the original trial**. If a clinician consented to audio recording, patients scheduled to see that clinician were also asked to allow audio recording. If a clinician did not consent to audio recording, patients scheduled to see that clinician were not asked about audio recording. When a clinician and patient both agreed to be recorded, a research assistant gave the patient a digital audio recorder before the visit started, showed the patient how to operate the recorder, and then retrieved the recorder after the visit. Clinicians were informed when a visit was being recorded.

2.2 Measures

2.2.1 Baseline measures—Patient baseline measures included demographics, the PHQ-9,[25, 26] the 12-Item Short Form Health Survey[27] (SF-12), six questions about patients' self-efficacy for communicating about mental health problems (rated on 5-point Likert scales),[28] and whether or not patients had several chronic medical conditions (e.g., arthritis, hypertension, diabetes). Clinician baseline measures included demographics, years in their current medical practice, and the number of half-days per week spent seeing patients.

2.2.2 Visit recording status—Investigators reviewed study consent documents to determine which patients and clinicians consented to be audio recorded. When both the patient and clinician consented to recording, we reviewed study data and tracking forms to determine whether that visit was actually recorded. Visits were counted as recorded if a digital audio file was present for the visit or if no audio file was present but study tracking forms explicitly indicated that the visit was recorded (i.e., the audio file was subsequently lost). Visits were counted as not recorded if no audio file was present and tracking forms gave no indication that the visit was recorded. Information about why visits were not recorded was abstracted when available.

2.2.3 Visit communication—After their visit, patients completed questionnaires about communication during that visit. Patients were asked whether two preventive health topics (i.e., diet and exercise habits) and six depressive symptoms (i.e., quantity and quality of sleep, mood, loss of interest in enjoyable activities, ability to perform usual activities, concentration, and thoughts of suicide or self-harm) were discussed or not. Patients were also asked whether the diagnosis of depression was discussed in any way and whether clinicians recommended any depression treatment. Clinicians completed post-visit questionnaires that included an estimate of visit length and 3 items rating visit difficulty (amount of time required, amount of effort required, and the degree to which the clinician found the visit difficult) on 3-point scales (less than average, about average, greater than average). Finally, clinicians indicated whether the patient in each visit was one of their established primary care patients.

2.3 Statistical analysis

2.3.1 Consent for recording—We examined the association between clinician baseline characteristics and clinician consent to be recorded using logistic regression. We started with a set of univariate (unadjusted) models, with clinician consent as the binary dependent variable and only one baseline characteristic at a time as a predictor. We then constructed a multivariable logistic regression model using as predictors all baseline characteristics with P-values < 0.2 in unadjusted analyses.[29, 30]

Analogous unadjusted and multivariable analyses were performed to evaluate factors associated with patient consent; we included only patients who were asked about audio recording and who saw a clinician who had consented to recording. Patient baseline SF-12 scores were converted to Mental Health Component Summary scores and Physical Health Component Summary scores.[31] Questions measuring self-efficacy for communicating

about mental health were summed to make a single rating scale (range 6–30). Patient-clinician sex concordance and race concordance were also examined as predictors of patient consent. Baseline characteristics with *P*-values < 0.2 in unadjusted analyses were considered for inclusion in the final model. We used hierarchical logistic regression and survey weights to account for patients being clustered within clinicians and for oversampling of patients with depressive symptoms, respectively. We also compared baseline patient characteristics by clinician consent status to investigate whether patients seeing clinicians who consented to recording differed systematically from patients seeing clinicians who did not consent.

2.3.2 Effect of recording—We performed a series of regression analyses to evaluate whether being recorded was associated with patient-clinician communication, treatment recommendations, or clinician burden. Separate analyses were conducted for 3 patient-reported dependent variables related to communication (number of depressive symptoms discussed, whether the diagnosis of depression was discussed, and whether diet or exercise was discussed), 1 patient-reported dependent variable related to treatment recommendations (whether the clinician recommended any depression treatment), and 2 clinician-reported dependent variables related to clinician burden (the sum of 3 items assessing overall visit difficulty (range 0–6) and estimated visit length). We analyzed communication about both depression and preventive health to investigate whether the effect of recording influenced communication about study-related and non-study related topics differently. Number of depressive symptoms, visit difficulty, and visit length were analyzed as continuous variables; all other variables were treated as binary **because of how they were measured in the original trial**. For each of the six dependent variables, we first examined associations with recording status in univariate analyses. We then built multivariable regression models to estimate the effect of being recorded on each dependent variable after adjusting for patient and clinician characteristics.

To verify that both the 3 items assessing visit difficulty and the 6 items measuring discussion of depression symptoms could be combined into single rating scales, we performed exploratory factor analysis for each set of items and examined factor loadings, eigenvalues, and scree plots. We also checked the assumptions required for creating summated rating scales by examining item-rest correlations and item-rest plots.[32] Analyses suggested that each set of items measured a single latent construct and could be summed to make a continuous variable.

Finally, we estimated the effect of being recorded on the six dependent variables discussed above by performing a propensity score analysis. The propensity score is a patient's probability of being recorded, conditional on observed covariates. Propensity scores are a method for removing bias in estimating the effect of an exposure (i.e., being recorded) on an outcome when using observational data.[33] We used multilevel statistical techniques to estimate each patient's probability of being recorded based on both patient- and clinician-specific characteristics.[34] We then used the inverse of the propensity score as a weight for each recorded patient and the inverse of one minus the propensity score as the weight for each non-recorded patient.[35, 36] **We used inverse probability of treatment weighting because recent simulation studies suggest that this method was one of two methods that eliminates systematic differences between treated and untreated subjects to a greater**

degree than stratification or covariate adjustment.[37, 38] We then performed a series of regression analyses parallel to those reported above to evaluate whether recording status was associated with any of six dependent variables. In these analyses, however, patients were weighted according to their propensity scores as described above. Due to the lack of overlap in the propensity scores for the two groups, we only included the 477 patients with propensity scores between 0.2 and 0.8 in these analyses. Further details of our propensity score analysis are available from the authors by request.

2.3.3 Missing data and regression diagnostics—Missing clinician baseline characteristics (5–8% of clinicians) were imputed first with publicly available data from clinic websites and the Medical Board of California and then, if necessary, with multiple imputation using the chained equation method.[39] Missing SF-12 scores (1.4% of patients) were imputed separately using the same statistical method. Clinician-reported visit difficulty and visit length contained 5–7% missing data; we did not impute these dependent variables. Model assumptions were checked by evaluating observed-expected tables (logistic regressions) and residual plots (linear regressions). All tests were two-sided with $\alpha = 0.05$. Propensity score analysis was performed using PROC GLIMMIX in SAS version 9.4. All other analyses were performed using Stata 13.1.

3. Results

The clinical trial enrolled 135 clinicians and 867 patients; the mean number of patients per clinician was 6.4 (range 1–20). Ninety-nine (73%) clinicians consented to be recorded. Four hundred eight (69%) patients who saw consenting clinicians agreed to be recorded. Eighty-two percent (*n*=334) of these visits were recorded. Tracking forms for the majority of non-recorded visits contained no information about why the visit was not recorded. Anecdotal reports suggested that the most common reasons were staff error (e.g., research assistant fails to recognize that a participant consented to be recorded) and technical failure (e.g., batteries die). Figure 1 shows the flow of the consent and recording process.

Table 1 compares characteristics of consenting and non-consenting clinicians. In unadjusted analyses, clinicians who consented to recording were significantly more likely to practice in academic primary care clinics, to spend fewer half-days per week seeing patients, to work in San Francisco, and to report longer tenures at their current practice location compared to clinicians who declined recording. In multivariable analyses (Table 2), working in academic primary care clinics was the strongest independent predictor of clinician consent to recording (OR 10.08, 95%CI 2.18 – 46.54, *P* = 0.003). The only other statistically significant predictor was years in current practice; clinicians who reported having worked longer in their current practice were more likely to consent to recording (OR 1.10, 95%CI 1.01 – 1.18, *P* = 0.02). We found no significant differences for any patient or visit-level variables when comparing patients seeing consenting clinicians to patients seeing non-consenting clinicians.

Table 3 compares characteristics of consenting and non-consenting patients among the 593 patients who were asked about recording and summarizes characteristics of the entire patient sample (*n* = 867). In univariate analyses, patients who consented to recording were

significantly more likely to be male, white, and to report better mental health compared to patients who declined recording. Patients were also more likely to consent if they lived in Sacramento or had arthritis, diabetes, or hypertension. In multivariable analysis (Table 4) only white race (OR 2.16, 95%CI 1.34 – 3.50, P = 0.002), having diabetes (OR 2.14, 95%CI 1.08 – 4.25, P = 0.03) and living in Sacramento (OR 1.82, 95%CI 1.13 – 2.94, P = 0.02) remained significantly associated with patient consent. Patient mental health status was not significantly associated with consent in multivariable analysis. In addition, patient consent was independent of clustering by clinician (i.e., the interclass correlation coefficient for clinician-level effects was 0). Exploratory analysis performed to investigate the unexpected finding related to diabetes revealed that patients with diabetes were significantly more likely to be identified as the clinician's established patient than were patients without diabetes.

Table 5 shows estimates of the effect of being recorded on visit communication, treatment recommendations, and clinician burden. In unadjusted analyses, patients whose visits were recorded reported discussing significantly more depressive symptoms and had significantly higher probabilities of discussing both the diagnosis of depression and at least one preventive health topic. None of these differences remained significant after controlling for other patient and visit-level characteristics. In both unadjusted and multivariable analyses, being recorded was not significantly associated with either clinician probability of recommending depression treatment or clinician burden. In exploratory analyses, there was no interaction between patients' baseline PHQ-9 scores and the effect of being recorded. We also performed a sensitivity analysis to explore whether the effect of recording remained non-significant when patient self-report variables (i.e., PHQ-9, SF-12, and self-efficacy) were omitted from the multivariable models. These results did not differ meaningfully from our primary analysis and so are not shown.

The propensity score analysis also indicated that being recorded had no significant effect on any of the tested dependent variables (Table 6). The predicted effect sizes associated with being recorded were similar to the effect sizes estimated from the multivariable analysis without propensity scores (Table 5).

4. Discussion and Conclusion

4.1. Discussion

In this study, we investigated clinician and patient characteristics associated with consent to audio record primary care visits and estimated the effect of audio recording on patient-clinician interactions using multivariable regression and propensity score analyses. Despite persistent worries that recording decreases the validity of research studies, we found few clinician or patient characteristics that were significantly associated with the odds of consenting to recording. Similarly, we found no evidence that recording introduced a significant Hawthorne effect that influenced patient-clinician communication, treatment recommendations, or clinician burden. If any effects of recording do exist, they are likely small and not clinically meaningful, at least for the kinds of visits and processes of care we examined.

Few prior studies have examined either clinician or patient factors associated with consent to recording. Most prior studies reported only univariate results and focused on attitudes towards video recording[6, 10] or the effect of discussion topic.[6, 11, 12, 15] **We did not evaluate attitudes towards recording, but our findings can inform future studies involving recording because we analyzed patient and clinician characteristics (e.g., demographics) that are commonly measured in health communication research**. Our study is one of the few to examine the effect of recording not only on communication but also on clinically important processes of care (i.e., treatment recommendations) and clinician burden.

Clinicians working in academic primary care clinics were significantly more likely to consent to recording than clinicians working in other settings. A recent review of recruitment strategies for research involving recording found that clinicians were more likely to consent if they were acquainted with study investigators.[9] This phenomenon likely contributed to increased clinician consent rates at academic clinics in our study. Clinicians who reported working longer in their present practice were also more likely to consent; these clinicians may be more efficient or more used to their environment and therefore, less worried that recording might affect clinic workflow. **The lack of significant associations with clinician consent for other clinic types and for clinician demographics may be due to insufficient statistical power rather than to absence of a meaningful association**.

Few patient characteristics were associated with the likelihood of patient consent. One notable exception was that white patients had nearly twice the odds of consenting relative to non-white patients (Table 4). Two prior studies examined associations between patient race and consent. One found no association but was underpowered.[13] The other found that non-white patients were significantly less likely to consent in unadjusted analyses, but that this difference became insignificant after controlling for patient attitudes about recording (e.g., concerns about confidentiality).[6] **We could not test this hypothesis because the original trial did not measure attitudes about recording**. Low rates of research participation for minority patients is a longstanding problem in both clinical and health communication research.[40] Patients with diabetes were more likely to consent to recording than were patients without diabetes. Possible explanations for this finding are that diabetic patients are more likely to see their established primary care clinician and/or have more frequent visits and so feel more comfortable in the clinic setting compared to patients without diabetes. Patients' mental health status was not significantly associated with patient consent in multivariable analysis. These findings are consistent with one previous study that also found patient mental health status was not significantly associated with consent for video recording in multivariable analyses.[11, 15]

We found no significant differences between recorded and non-recorded visits for patient-clinician communication, depression treatment recommendations, or clinician burden in either multivariable analyses or propensity score analyses. In the context of prior literature, these findings suggest that while participating in a communication study may have significant effects on patient-clinician communication, adding the element of recording does not introduce significant additional effects. Our findings are consistent with prior studies

that found little evidence to suggest that video cameras influence verbal or non-verbal behaviors during visits[41, 42] and no differences between video recorded and non-recorded visits in terms of visit content,[43] patient satisfaction,[18] or patient arousal.[44] On the other hand, two studies involving recording (or direct observation) that compared participants and non-participants found that participation was associated with significant changes in antibiotic prescribing[45] and the quality of primary care visits.[46] The findings in these two studies were likely due to awareness of study participation, knowledge of study goals, and/or information bias[8] rather than to being recorded.

Our study has several limitations. Consent for recording was optional in the parent study, so our findings may not generalize to studies in which audio recording is required for participation. However, this approach allowed collection of detailed data about non-recorded visits and thus avoided potential information bias due to comparing participants and non-participants. Our estimates of the effect of recording are based on observational data; however, our results did not change when we performed propensity score analyses or excluded self-reported patient measures from multivariable analyses. A randomized trial examining the effects of recording on patient-clinician communication would likely require covert recording, which institutional review boards might hesitate to allow. Our study only measured the content of communication (i.e., topics of discussion) and so did not evaluate whether being recorded was associated with changes in more subtle communication processes (e.g., increased hesitancy when discussing depression). Finally, our measures of communication were patient-reported, which introduces the possibility of measurement error.[47] **However, use of patient-reported outcomes would not affect our conclusions about the validity of research involving recording unless recording had different effects on patient behavior during visits compared to patient recall after visits. These limitations** are unavoidable without covert recording and are unlikely to obscure clinically meaningful effects on communication or treatment recommendations.

4.2 Conclusion

Our analysis of data from a clinical trial to engage primary care patients in communication about depression found that audio recording visits did not introduce significant threats to research validity. We found little evidence that recording introduced selection bias or had significant effects on either communication or processes of care. Patient and clinician concerns about recording must be taken seriously, but most stakeholders today are amenable to recording primary care visits for research or quality improvement.[7] As patients and health care payers make increasing demands for high-quality care, audio or video recording of clinical interactions for research and quality improvement may become routine. In the meantime, findings from this study should help establish confidence in the validity of research involving recorded clinical interactions.

4.3 Practice implications

If corroborated by future research, our findings suggest that the benefits of recording patient-clinician interactions outweigh the potential risks of bias in most situations. Recording interactions is a powerful method for understanding the content of clinic visits[5] and has substantial potential for assessing and improving health care quality,[48] training

clinicians,[49] and empowering patients.[50] Our results should encourage health communication researchers to integrate audio recording into study protocols without worrying that doing so will reduce the validity of their results. Of course, researchers should also take steps to ensure that non-white patients and clinicians outside of academic settings are adequately represented in study samples. Finally, our analysis demonstrates that incorporating optional consent for recording into future studies is a useful approach for distinguishing the effects of recording from the effects of study participation in general,[51] and for exploring whether our findings generalize to other contexts.

ACKNOWLEDGEMENTS

REFERENCES

1. Noordman J, van der Weijden T, van Dulmen S. Communication-related behavior change techniques used in face-to-face lifestyle interventions in primary care: A systematic review of the literature. Patient Educ Couns. 2012; 89:227–244. [PubMed: 22878028]

2. Haskard-Zolnierek K, DiMatteo MR. Physician communication and patient adherence to treatment: A meta-analysis. Med Care. 2009; 47:826–834. [PubMed: 19584762]

3. Trangle, M.; Dieperink, B.; Gabert, T., et al. Major Depression in Adults in Primary Care. Bloomington, MN: Institute for Clinical Systems Improvement; 2012.

4. Donabedian A. The quality of care: How can It be assessed? J Am Med Assoc. 1988; 260:1743–1748.

5. Hrisos S, Eccles MP, Francis JJ, et al. Are there valid proxy measures of clinical behaviour? a systematic review. Implementation science : IS. 2009; 4:37. [PubMed: 19575790]

6. Bradshaw D, Cohen CE, Day S, Mandalia S, Theobald N. Acceptability of video recordings of consultations in HIV and genitourinary medicine (GUM). Patient Educ Couns. 2013; 92:279–280. [PubMed: 23664233]

7. Rushmer R, Themessel-Huber M, Coyle J, Humphris G, Dowell J, Williams B. Is the routine recording of primary care consultations possible … and desirable? Lessons for researchers from a consultation with multiple stakeholders. Patient Educ Couns. 2011; 82:247–253. [PubMed: 20547028]

8. Grimes DA, Schulz KF. Bias and causal associations in observational research. Lancet. 2002; 359:248–252. [PubMed: 11812579]

9. Themessl-Huber M, Humphris G, Dowell J, Macgillivray S, Rushmer R, Williams B. Audio-visual recording of patient-GP consultations for research purposes: A literature review on recruiting rates and strategies. Patient Educ Couns. 2008; 71:157–168. [PubMed: 18356003]

10. Coleman T. Sampling for qualitative research using quantitative methods. 2. Characteristics of GPs who agree to video-taping of consultations. Fam Pract. 1996; 13:531–535. [PubMed: 9023530]

11. Howe A. Refusal of videorecording: what factors may influence patient consent? Fam Pract. 1997; 14:233–237. [PubMed: 9201498]

12. Martin E, Martin PM. The reactions of patients to a video camera in the consulting room. J R Coll Gen Pract. 1984; 34:607–610. [PubMed: 6502570]

13. Neal RD, Ali N, Allgar V, Coleman T. Consent rates for video-recording general practice consultations: effect of ethnicity and other factors. Fam Pract. 2004; 21:219–220. [PubMed: 15020395]

14. van den Brink-Muinen, A.; Verhaak, PFM.; Bensing, J., et al. The Eurocommunication study. An international comparitive study in six European countries on doctor-patient communication. Utrecht, Netherlands: NIVEL; 1999.

15. Coleman T, Manku-Scott T. Comparison of video-recorded consultations with those in which patients' consent is withheld. Br J Gen Pract. 1998; 48:971–974. [PubMed: 9624767]

16. McCambridge J, Kypri K, Elbourne D. Research participation effects: a skeleton in the methodological cupboard. J Clin Epidemiol. 2014; 67:845–849. [PubMed: 24766858]

17. Merrett F. Reflections on the Hawthorne Effect. Educational Psychology. 2006; 26:143–146.

18. Campbell LM, Sullivan F, Murray TS. Videotaping of general practice consultations: effect on patient satisfaction. Brit Med J. 1995; 311:236. [PubMed: 7627043]

19. Ram P, Grol R, Rethans JJ, Schouten B, van der Vleuten C, Kester A. Assessment of general practitioners by video observation of communicative and medical performance in daily practice: issues of validity, reliability and feasibility. Med Educ. 1999; 33:447–454. [PubMed: 10354322]

20. Herzmark G. Reactions of patients to video recording of consultations in general practice. Br Med J (Clin Res Ed). 1985; 291:315–317.

21. Hargreaves PN, Peppiatt R. Is videotaping of consultations acceptable to patients attending a hospice day centre? Palliat Med. 2001; 15:49–54. [PubMed: 11212468]

22. Kravitz RL, Franks P, Feldman MD, et al. Patient engagement programs for recognition and initial treatment of depression in primary care: a randomized trial. J Amer Med Assoc. 2013; 310:1818–1828.

23. Kroenke K, Strine TW, Spitzer RL, Williams JBW, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. J Affect Disorders. 2009; 114:163–173. [PubMed: 18752852]

24. Tancredi DJ, Slee CK, Jerant A, et al. Targeted versus tailored multimedia patient engagement to enhance depression recognition and treatment in primary care: randomized controlled trial protocol for the AMEP2 study. BMC Health Serv Res. 2013; 13:141. [PubMed: 23594572]

25. Kroenke K, Spitzer RL. The PHQ-9: A new depression diagnostic and severity measure. Psychiat Ann. 2002; 32:509–515.

26. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med. 2001; 16:606–613. [PubMed: 11556941]

27. Ware J Jr. Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care. 1996; 34:220–233. [PubMed: 8628042]

28. Maly RC, Frank JC, Marshall GN, DiMatteo MR, Reuben DB. Perceived efficacy in patient-physician interactions (PEPPI): validation of an instrument in older persons. J Am Geriatr Soc. 1998; 46:889–894. [PubMed: 9670878]

29. Miller, AJ. Subset selection in regression. 2nd ed. Boca Raton: Chapman & Hall/CRC; 2002. p. 42-43.

30. Hosmer, DW.; Lemeshow, S. Applied logistic regression. 2nd ed. New York: Wiley; 2000. p. 91-97.

31. Fleishman JA, Selim AJ, Kazis LE. Deriving SF-12v2 physical and mental health summary scores: a comparison of different scoring algorithms. Qual Life Res. 2010; 19:231–241. [PubMed: 20094805]

32. McIver, JP.; Carmines, EG. Unidimensional scaling. Beverly Hills: Sage Publications; 1981.

33. D'Agostino RB Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. Stat Med. 1998; 17:2265–2281. [PubMed: 9802183]

34. Griswold ME, Localio AR, Mulrow C. Propensity score adjustment with multilevel data: setting your sites on decreasing selection bias. Ann Intern Med. 2010; 152:393–395. [PubMed: 20231571]

35. Rosenbaum PR, Rubin DB. The Central Role of the Propensity Score in Observational Studies for Causal Effects. Biometrika. 1983; 70:41–55.

36. Lunceford JK, Davidian M. Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. Stat Med. 2004; 23:2937–2960. [PubMed: 15351954]

37. Austin PC. The relative ability of different propensity score methods to balance measured covariates between treated and untreated subjects in observational studies. Med Decis Making. 2009; 29:661–677. [PubMed: 19684288]

38. Austin PC, Grootendorst P, Anderson GM. A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a Monte Carlo study. Stat Med. 2007; 26:734–753. [PubMed: 16708349]

39. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Stat Med. 2011; 30:377–399. [PubMed: 21225900]

40. Cooper, LA.; Roter, DL. Patient-provider communication: The effect of race and ethnicity on process and outcomes of healthcare. In: Smedley, BD.; Stith, AY.; Nelson, AR., editors. Institute of Medicine (U.S.). Committee on Understanding and Eliminating Racial and Ethnic Disparities in Health Care., Unequal treatment: Confronting racial and ethnic disparities in health care. Washington, D.C.: National Academy Press; 2003. p. 552-593.

41. Penner LA, Orom H, Albrecht TL, Franks MM, Foster TS, Ruckdeschel JC. Camera-related behaviors during video recorded medical interactions. J Nonverbal Behav. 2007; 31:99–117.

42. Zhou Y, Forbes GM, Humphris G. Camera-related behaviours of female dental nurses and nursery school children during fluoride varnish application interactions in nursery school settings. Int J Paediatr Dent. 2010; 20:374–381. [PubMed: 20545791]

43. Pringle M, Stewart-Evans C. Does awareness of being video recorded affect doctors consultation behavior? Brit J Gen Pract. 1990; 40:455–458. [PubMed: 2271278]

44. Pringle M, Robins S, Brown G. Assessing the consultation: methods of observing trainees in general practice. Br Med J (Clin Res Ed). 1984; 288:1659–1660.

45. Mangione-Smith R, Elliott MN, McDonald L, McGlynn EA. An observational study of antibiotic prescribing behavior and the Hawthorne effect. Health Serv Res. 2002; 37:1603–1623. [PubMed: 12546288]

46. Leonard K, Masatu MC. Outpatient process quality evaluation and the Hawthorne Effect. Soc Sci Med. 2006; 63:2330–2340. [PubMed: 16887245]

47. Henry SG, Feng B, Franks P, et al. Methods for Assessing Patient-Clinician Communication about Depression in Primary Care: What You See Depends on How You Look. Health Serv Res. 2014; 49:1684–1700. [PubMed: 24837881]

48. Makary MA. The power of video recording: taking quality to the next level. J Amer Med Assoc. 2013; 309:1591–1592.

49. Henry SG, Holmboe ES, Frankel RM. Evidence-based competencies for improving communication skills in graduate medical education: a review with suggestions for implementation. Med Teach. 2013; 35:395–403. [PubMed: 23444891]

50. Tsulukidze M, Durand MA, Barr PJ, Mead T, Elwyn G. Providing recording of clinical consultation to patients - a highly valued but underutilized intervention: a scoping review. Patient Educ Couns. 2014; 95:297–304. [PubMed: 24630697]

51. McCambridge J, Witton J, Elbourne DR. Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. J Clin Epidemiol. 2014; 67:267–277. [PubMed: 24275499]

Highlights

- We investigate factors associated with consent to audio record clinic visits

- We examined the effect of recording on patient-clinician interactions

- Few patient or clinician factors were associated with likelihood of consent

- We found no significant effects of recording on patient-clinician interactions
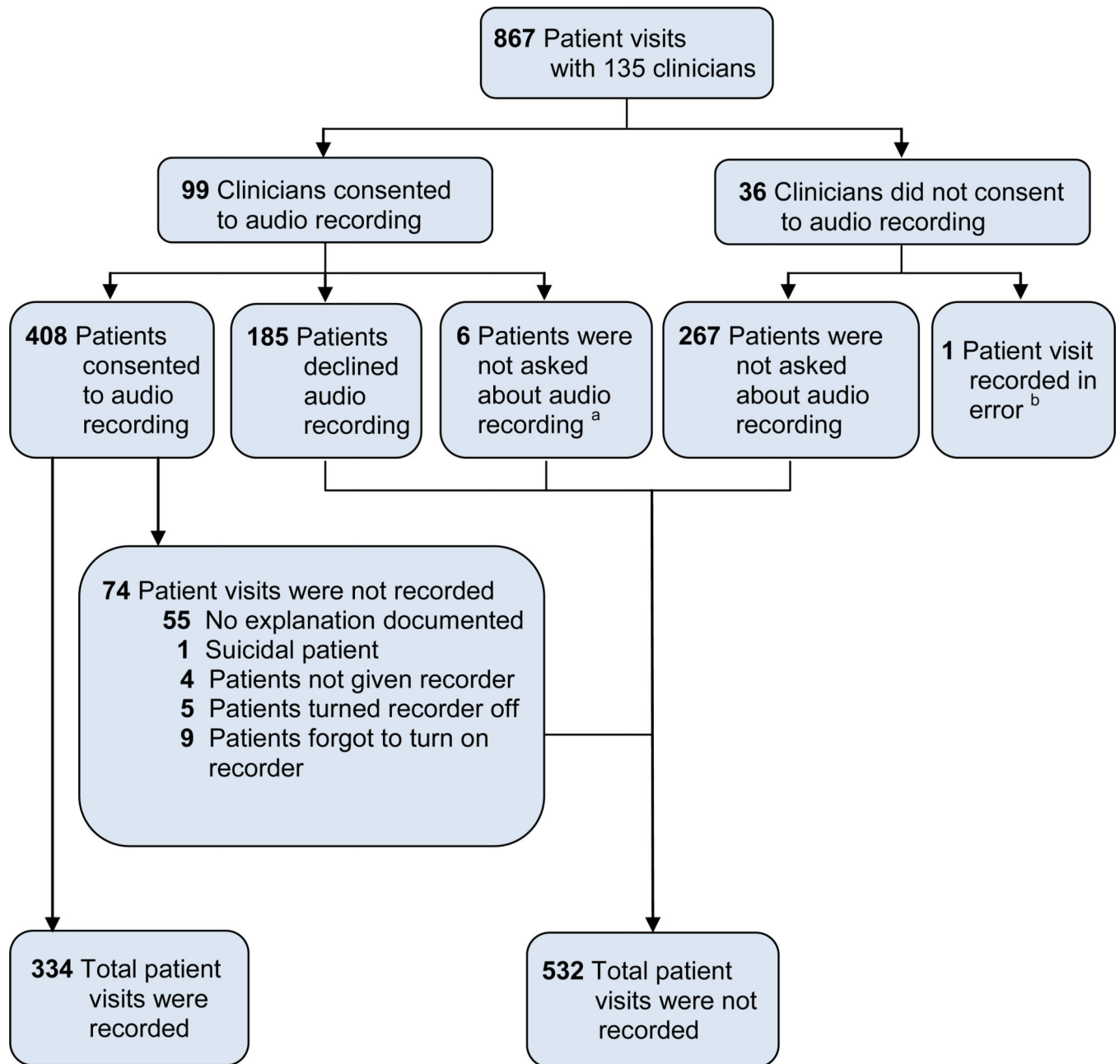
Figure 1. Flow of Study Consent and Recording Process

[a] Patients were not asked about recording due to logistical constraints

[b] Patient was asked about and consented to recording; however, the clinician had previously declined recording. The clinician was aware that the visit was being recorded. This patient is excluded from statistical comparisons in this article.

Table 1

Comparison of clinician characteristics by consent status[a]

	Declined recording (n=36)		Consented to recording (n=99)		P value
Age (mean, SD)	42.7	9.3	44.7	8.97	0.28
Female sex (no, %)	24	66.7%	50	50.5%	0.10
Race (no, %)					
Non-Hispanic white	13	38.2%	56	59.6%	0.08
Non-Hispanic black	1	2.9%	1	1.1%	
Hispanic	2	5.9%	7	7.5%	
Asian / Pacific islander	15	44.1%	28	29.8%	
Other / mixed race	3	8.8%	2	2.1%	
Specialty (no, %)					
Family practice	17	47.2%	32	32.3%	0.11
Internal medicine	19	52.8%	67	67.7%	
Clinician type (no, %)					
Physician	36	100.0%	93	93.9%	0.34
Nurse practitioner	0	0.0%	6	6.1%	
Clinic type (no, %)					
Community primary care	23	63.9%	31	31.3%	0.001
Academic primary care	4	11.1%	39	39.4%	
Veterans Affairs	8	22.2%	20	20.2%	
Urgent care	1	2.8%	9	9.1%	
Geographic area (no, %)					
Sacramento	29	80.6%	51	51.5%	0.004
San Francisco	7	19.4%	48	48.5%	
Years since completing training (mean, SD)	11.5	9.5	14.4	8.9	0.10
Years in current practice site (mean, SD)	7.4	6.3	10.5	6.9	0.02
Half-days of clinic per week (mean, SD)	6.4	2.8	4.9	2.8	0.01

	Declined recording (n=36)	Consented to recording (n=99)	P value
Number of patients per half-day (mean, SD)	10.5 4.1	9.8 2.8	0.27

[a] Groups were compared using the two-sample t-test for continuous variables, Fisher's exact test for categorical variables and clinician type, and logistic regression for binary variables other than clinician type. Not all categorical variables add up to column totals due to missing data.

Table 2

Factors associated with clinician consent to be recorded[a]

	OR	95%CI	AOR[b]	95%CI
Female sex	0.51	0.23 – 1.13	0.42	0.17 – 1.05
White race	2.38	1.06 – 5.32	1.63	0.65 – 4.10
Half-days of clinic per week	0.83	0.72 – 0.96	1.00	0.81 – 1.24
Years in current practice	1.08	1.01 – 1.16	1.10	1.01 – 1.18
Clinic type				
Community primary care	ref	–	ref	–
Academic primary care	7.23	2.26 – 23.12	10.08	2.18 – 46.54
Veterans affairs	1.85	0.70 – 4.95	2.32	0.67 – 8.04
Urgent care	6.68	0.79 – 56.48	8.72	0.88 – 86.14

[a]Odds ratios indicate odds of clinician consenting to be recorded.

[b]Adjusted for all listed independent variables. Multiple imputation was used to account for missing data. Geographic area and specialty were not included in the final model because they were collinear with clinic type. Years in practice was not included because it was collinear with years in current practice.

Table 3

Comparison of patient characteristics by consent status

	Declined recording (n=185)		Consented to recording (n=408)		P value[a]	All patients (n=867)	
Age (mean, SD)	50.1	11.8	52.1	12.0	0.06	51.8	11.8
Female sex (no, %)	113	61.1%	213	52.2%	0.045	486	56.1%
Race (no, %)							
Non-Hispanic white	85	46.0%	251	61.5%	0.002	486	56.1%
Non-Hispanic black	41	22.2%	62	15.2%		155	17.9%
Hispanic	33	17.8%	47	11.5%		123	14.2%
Asian / Pacific islander	21	11.4%	28	6.9%		71	8.2%
Other / mixed race	5	2.7%	20	4.9%		32	3.7%
Highest education (no, %)							
No high school degree	8	4.4%	10	2.5%	0.41	25	2.9%
High school graduate	19	10.4%	57	14.1%		129	15.0%
Some college	70	38.3%	154	38.1%		316	36.7%
College graduate	86	47.0%	183	45.3%		391	45.4%
Household income (no, %)							
< $35,000 / year	77	41.6%	149	36.5%	0.49	317	36.6%
$35,000 – 75,000 /year	48	26.0%	113	27.7%		244	28.1%
$75,000 / year	60	32.4%	146	35.8%		306	35.3%
Employment status (no, %)							
Works for wages	75	41.0%	176	43.6%	0.42	372	42.9%
Self employed	12	6.6%	24	5.9%		57	6.6%
Out of work	25	13.7%	40	9.9%		88	10.2%
Homemaker	8	4.4%	8	2.0%		25	2.9%
Student	2	1.1%	2	0.5%		8	0.9%
Retired	31	16.9%	82	20.3%		171	19.7%
Unable to work	30	16.4%	72	17.8%		140	16.2%
Clinic type (no, %)							
Community primary care	73	39.5%	177	43.4%	0.34	449	51.8%

	Declined recording (n=185)		Consented to recording (n=408)		P value[a]	All patients (n=867)	
Academic primary care	63	34.1%	133	32.6%		214	24.7%
Veterans Affairs	25	13.5%	63	15.4%		130	15.0%
Urgent care	24	13.0%	35	8.6%		74	8.5%
Geographic area (no, %)							
San Francisco	84	45.4%	120	29.4%	<0.001	237	27.3%
Sacramento	101	54.6%	288	70.6%		630	72.7%
Sex concordant visit (no, %)	123	66.5%	262	64.2%	0.59	564	65.1%
Race concordant visit[b](no, %)	58	32.4%	154	39.2%	0.12	291	33.6%
PHQ-9 score[c](mean, SD)	7.9	5.9	7.4	5.7	0.31	7.4	5.7
SF-12 PCS[d](mean, SD)	42.0	13.8	40.9	13.5	0.37	41.5	13.5
SF-12 MCS[d](mean, SD)	44.2	12.7	47.0	11.9	0.01	46.3	12.1
Self-efficacy for communicating about mental health (mean, SD)[e]	21.4	6.3	22.2	5.6	0.13	21.8	5.8
Arthritis[f](no, %)	58	31.4%	171	41.9%	0.02	339	39.1%
Hypertension[f](no, %)	62	33.5%	176	43.1%	0.03	359	41.4%
Diabetes[f](no, %)	23	12.4%	86	21.1%	0.01	152	17.5%

[a]P-values refer to comparisons between the 185 patients who declined and the 408 patients who consented to recording. Groups were compared using the two-sample t-test for continuous variables, logistic regression for binary variables, and Pearson's chi-squared test for categorical variables. Not all categorical variables add up to column totals due to missing data.

[b]Race concordant visits were defined as visits in which both patient and clinician listed the same race from among the 5 categories in this table.

[c]PHQ-9 indicates the Personal Health Questionnaire-9; scores range from 0 to 30 with higher scores indicating more severe depressive symptoms.

[d]SF-12 PCS and SF-12 MCS indicate the physical health component summary score and the mental health component summary score of the 12-Item Short Form Health Survey. Both scores range from 0 to 100; higher values indicate better health.

[e]Range 0–30; higher values indicate greater self-efficacy.

[f]Measured as present or absent by baseline patient report

Table 4

Factors associated with patient consent to be recorded[a]

	OR	95%CI	AOR[b]	95%CI
Age	1.01	1.00 – 1.03	1.00	0.98 – 1.03
Female sex	0.70	0.49 – 0.99	0.92	0.57 – 1.49
White race	1.88	1.32 – 2.67	2.16	1.34 – 3.50
Sacramento[c]	2.00	1.39 – 2.86	1.82	1.13 – 2.94
SF-12 MCS[d]	1.02	1.00 – 1.03	1.01	0.99 – 1.03
Arthritis[e]	1.58	1.09 – 2.28	1.38	0.85 – 2.25
Hypertension[e]	1.51	1.05 – 2.16	1.29	0.74 – 2.26
Diabetes[e]	1.88	1.14 – 3.09	2.14	1.08 – 4.25
Self-efficacy[f]	1.02	0.99 – 1.05	1.01	0.97 – 1.05

[a]Odds ratios indicate odds of patients consenting.

[b]Adjusted for all listed values in addition to randomization arm and for oversampling of patients with depressive symptoms. Race concordance was not included in the final model because it was collinear with patient race. Multiple imputation was used to account for 12 missing values in SF-12 MCS. Estimates are not adjusted for clustering within clinician because clustering did not explain any of the observed variance (i.e., interclass correlation coefficient for clinician level effects was 0).

[c]Reference group is patients getting care in San Francisco.

[d]SF-12 MCS indicates the mental health component summary score of the 12-Item Short Form Health Survey. Scale ranges from 0 to 100; higher values indicate better health.

[e]Measured as present or absent by patient report

[f]Patient self-efficacy for communicating about mental health; range 0–30 with higher values indicating greater self-efficacy.

Table 5

Effect of being recorded on communication, treatment recommendations, and clinician burden[a]

Communication	Unadjusted analysis			Multivariable analysis[b]		
	Recorded	Not recorded	P-value	Recorded	Not Recorded	P-value
Number of depressive symptoms discussed[c]	2.9	2.5	0.003	2.7	2.5	0.19
Probability that either diet or exercise was discussed	0.74	0.67	0.03	0.72	0.69	0.49
Probability that the diagnosis of depression was discussed	0.60	0.53	0.04	0.59	0.55	0.35
Treatment recommendation						
Probability that the clinician recommended any depression treatment	0.20	0.17	0.26	0.17	0.15	0.43
Clinician-reported visit burden						
Rating of visit difficulty[d]	3.3	3.2	0.31	3.3	3.2	0.47
Visit length (minutes)	24.6	24.0	0.34	24.0	24.1	0.89

[a] Columns show predicted results for recorded versus non-recorded visits based on the underlying regression models.

[b] Multivariable analyses are adjusted for patient age, sex, race, clinic site, randomization assignment, baseline PHQ-9 score, geographic region, SF-12 physical health component score, clustering by clinician, and (for communication variables only) patient self-efficacy for communicating about mental health. Multiple imputation was used to account for 12 missing values in SF-12 physical health component score.

[c] Range 0–6

[d] Rated from 0 to 6; higher numbers indicate greater difficulty.

Table 6

Propensity score analysis for the effect of being recorded on communication, treatment recommendations, and clinician burden

	Unadjusted analysis[a]		Multivariable analysis[b]			
	Model coefficient (*SE*) for recorded versus non-recorded visits	*P*-value	Model coefficient (*SE*) for recorded versus non-recorded visits	*P*-value	Estimated difference (*SE*) in outcome for recorded versus non-recorded visits[c]	*P*-value
Communication processes						
Number of depressive symptoms discussed[d]	0.09 (0.18)	0.59	0.08 (0.18)	0.64	0.18 (0.14)	0.20
Either diet or exercise was discussed	0.02 (0.20)	0.93	0.01 (0.22)	0.95	0.12 (0.17)	0.49
Diagnosis of depression was discussed	0.07 (0.18)	0.70	0.06 (0.19)	0.75	0.16 (0.16)	0.34
Treatment recommendation						
Clinician recommended any depression treatment	0.34 (0.25)	0.17	0.36 (0.26)	0.17	0.19 (0.20)	0.33
Clinician-reported visit burden						
Rating of visit difficulty[e]	0.10 (0.13)	0.44	0.09 (0.13)	0.48	0.09 (0.11)	0.41
Visit length (minutes)	0.26 (0.68)	0.71	0.20 (0.67)	0.77	0.03 (0.64)	0.96

[a] Unadjusted analyses are controlled only for clustering by clinician using a random effect.

[b] Multivariable analyses are adjusted for patient age, sex, race, clinic site, randomization assignment, baseline PHQ-9 score, geographic area, SF-12 physical health component score, clustering by clinician, and (for communication process variables only) self-efficacy for communicating about mental health.

[c] **Results in this column are for the entire sample (*n*= 866) to facilitate comparison with Table 5. Results in the first 2 columns reflect only those visits with propensity scores between 0.2 and 0.8 (*n*= 477).**

[d] Range 0–6

[e] Rated from 0 to 6; higher numbers indicate greater difficulty.